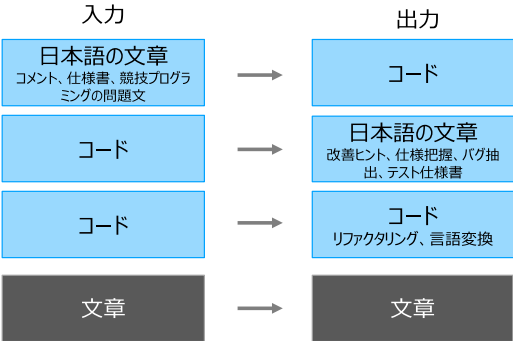


日本語とソフトウェア開発に特化した基盤モデルの構築・事業成果概要

実施者 フューチャー株式会社

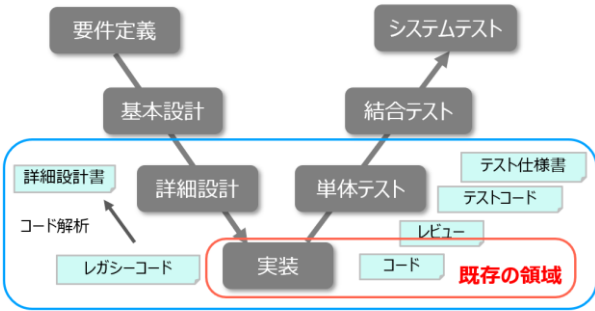
事業概要 ソフトウェア開発の省力化、高品質化実現に向けて、Llama 3.1を継続事前学習することにより、日本語とソフトウェア開発に特化した基盤モデルを開発する。

汎用モデル Llama 3.1 8B をソースコード、ソフトウェア開発に関連した日本語データ等を用いて領域適応することにより、日本語ソフトウェア開発に特化した基盤モデルの構築を行う。これにより、日本語からのコード生成や、コードからのレビュー生成等において従来モデルを上回る精度を達成することを目指し、ソフトウェア開発の省力化、高品質化を目指す。



社会実装イメージ

本事業で開発した基盤モデルを用いて、ソフトウェア開発支援ツールの構築に取り組む。本ツールでは、従来のコード補完だけでなく、設計やテスト、レビュー等の領域まで対応領域を拡張することを目指す。



事業成果

Llama 3.1 8Bをベースとし、日本語・英語・ソースコード合わせて約300Bトークンの継続事前学習を行い、その後ソフトウェア開発に関する合成指示データ500万件を用いて事後学習を行った。定量的な性能計測のため、国際的に広く用いられているコード補完ベンチマークデータである“HumanEval” およびそれを日本語訳した“JHumanEval”を中心にコード補完性能を計測した。また、HumanEvalのテストセットを拡張した“HuamnEval+”、コードの途中補完性能を測る“SantaCoder-FIM”についてもLlama 3.1と比較を行った。開発したモデルは上記全てのベンチマークにおいて、Llama 3.1 8Bを超える性能を達成し、一部タスクについてはLlama 3.1 70Bを超える性能が得られた。今後、当社独自のソリューションに本モデルを導入し、ソフトウェア開発の省力化、高品質化が達成できることを確認するとともに、モデル、評価データおよびノウハウの公開を進める。

評価データ	Llama 3.1 8B (ベースライン)	GENIACモデル
HumanEval	0.6311	0.6835 (+0.0524)
JHumanEval	0.5061	0.6335 (+0.1274)
HumanEval+	0.5872	0.6360 (+0.0488)
SantaCoder-FIM (Python)	0.4468	0.5139 (+0.0671)
SantaCoder-FIM (Java)	0.3506	0.5478 (+0.1972)